

WHAT IS CLAIMED IS:

1. A method of generating a virtual suffix tree (ViST) structure for searching XML documents, comprising:

receiving one or more XML documents;

5 converting the one or more XML documents into one or more structure-encoded sequences; and

generating the ViST structure comprising:

generating a D-Ancestor index;

generating an S-Ancestor index; and

10 generating a doc-ID index.

2. The method of claim 1, wherein generating a D-Ancestor index comprises generating a D-Ancestor B<sup>+</sup>Tree, wherein the D-Ancestor B<sup>+</sup>Tree indexes one or more (key, data) pairs and wherein the key element is a unique (symbol,path) pair in the one or  
15 more structure-encoded sequences, and the data element is a pointer to an S-Ancestor B<sup>+</sup>Tree.

3. The method of claim 1, wherein generating an S-Ancestor index comprises generating an S-Ancestor B<sup>+</sup>Tree, wherein the S-Ancestor B<sup>+</sup>Tree indexes one or more  
20 keys and wherein each of the one or more keys is a pair [begin-ID,end-ID].

4. The method of claim 3, wherein generating an S-Ancestor B<sup>+</sup>Tree, wherein the S-Ancestor B<sup>+</sup>Tree indexes one or more keys and wherein each of the one or more keys is a pair [begin-ID,end-ID] comprises generating an S-Ancestor index comprises  
25 generating an S-Ancestor B<sup>+</sup>Tree, wherein the S-Ancestor B<sup>+</sup>Tree indexes one or more keys and wherein each of the one or more keys is a pair (begin-ID,end-ID), wherein IDs of descendent nodes of a node whose label is (begin-ID,end-ID) are in the range of [begin-ID,end-ID].

30

5. The method of claim 1, wherein generating a doc-ID index comprises generating a doc-ID B<sup>+</sup>Tree, wherein the doc-ID B<sup>+</sup>Tree indexes one or more (key,data) pairs and wherein the key element is a node ID, and the data element is a list of XML document IDs.

6. A method of answering an XML query, comprising:  
receiving an XML query;  
transforming the XML query into a structure-encoded sequence;  
searching a ViST structure using the structure-encoded sequence and  
returning one or more document IDs.

7. The method of claim 6, wherein searching a ViST structure using the structure encoded sequence, comprises:

(a). assuming the query sequence is  $\langle q_1, q_2, \dots, q_n \rangle$ ;

(b) assigning  $i=1, \text{begin}=0, \text{end}=\text{infinity}$ ;

(c) searching a D-Ancestor B<sup>+</sup>Tree using key  $q_i$ , which returns an S-Ancestor B<sup>+</sup>Tree; and

(d) performing a range search (begin,end) on the S-Ancestor B<sup>+</sup>Tree, wherein performing the range search comprises:

(e) returning a set of ranges  $(x_1, y_1), \dots, (x_n, y_n)$ ;

(f) for each  $(x_i, y_i)$  doing (g) and (h);

(g) if  $(i=n)$  then

performing a range query  $(x_i, y_i)$  on the doc-ID index and returning one or more document IDs;

(h) if  $(i < n)$  then

assigning  $i=i+1$ ;  $\text{begin}=x_i$ ,  $\text{end}=y_i$ ;

going to (c).

8. A method of dynamically updating the ViST structure, comprising receiving a new XML document;

transforming the XML document into a structure-encoded sequence;  
inserting each element of the sequence into D-Ancestor B<sup>+</sup>Tree;  
assigning a new label if the step of inserting creates a new node; and  
inserting the new label into the S-Ancestor B<sup>+</sup>Tree.

5

9. The method of claim 8, wherein assigning a new label if the step of inserting creates a new node comprises assigning a new label (x,y) if the step of inserting creates a new node.

10

10. The method of claim 8, wherein inserting the new label into the S-Ancestor B<sup>+</sup>Tree comprises inserting the new label (x,y) into the S-Ancestor B<sup>+</sup>Tree.

15

11. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method of generating a virtual suffix tree (ViST) structure for searching XML documents, comprising the steps of:

receiving one or more XML documents;  
converting the one or more XML documents into one or more  
structure-encoded sequences;  
generating the ViST structure comprising:  
generating a D-Ancestor index;  
generating an S-Ancestor index; and  
generating a doc-ID index.

20

25

12. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method answering an XML query, comprising the steps of:

receiving an XML query;  
transforming the XML query into a structure-encoded sequence;  
searching a ViST structure using the structure-encoded sequence and  
returning one or more document IDs.

30

13. A machine-readable medium having instructions stored thereon for execution by a processor to perform a method of dynamically updating the ViST structure, comprising the steps of:

- 5 receiving a new XML document
- transforming the XML document into a structure-encoded sequence
- inserting each element of the sequence into D-Ancestor B<sup>+</sup>Tree;
- assigning a new label if the step of inserting creates a new node; and
- 10 inserting the new label into the S-Ancestor B<sup>+</sup>Tree.